

WHAT IS CLAIMED IS:

1. A document information processing apparatus comprising: a plain document input unit for inputting a plain document;
5 document;

a dictionary storage unit for storing a dictionary used for form element analysis and syntactic analysis;

a form element analyzer for performing a form element analysis on the plain document inputted from said plain document input unit by using the dictionary stored in said dictionary storage unit so as to decompose the plain document into tokens;

a syntax analyzer for analyzing a part of speech of each of the tokens obtained by said form element analyzer based on a syntax of said plain document so as to generate a structured document containing meaningful words;

a data storage unit for storing data used for a markup process;

an element refinement processing unit for performing the markup process of reading and adding data associated with each of the meaningful words included in the structured document generated by said syntax analyzer and stored in said data storage unit to each of the meaningful words so as to generate a markup document; and

a markup document output unit for outputting the markup document generated by said element refinement processing unit.

2. The document information processing apparatus according to claim 1, wherein said apparatus further includes a text document input unit for generating a plain text document and sending it to said plain document input unit.

3. The document information processing apparatus according to claim 1, wherein said apparatus further includes an external communication input unit for inputting a plain document from outside said apparatus by communication, and sends it to said plain document input unit.

4. The document information processing apparatus according to claim 1, wherein said apparatus further includes an external communication input unit for inputting a document on which a markup process is performed from outside said apparatus by communication, and a markup document analyzer for analyzing the document on which the markup process is performed from said external communication input unit and removing markups from the document so as to generate a plain document and send it to said markup document input unit.

5. The document information processing apparatus according to Claim 1, wherein said apparatus further includes an external communication input unit for inputting an e-mail from outside said apparatus by communication, and an e-mail structure analyzer for analyzing a structure of the e-mail from said external communication input unit so as to generate a plain document, and for sending the plain document to a markup document input unit.

6. The document information processing apparatus Claim 2, wherein said apparatus further includes an external communication output unit for sending the markup document from said markup document output unit to outside said apparatus.

7. The document information processing apparatus Claim 3, wherein said apparatus further includes an external communication output unit for sending the markup document from said markup document output unit to outside said apparatus.

8. The document information processing apparatus Claim 4, wherein said apparatus further includes an external communication output unit for sending the markup document from said markup document output unit to outside said apparatus.

9. The document information processing apparatus Claim 5, wherein said apparatus further includes an external communication output unit for sending the markup document from said markup document output unit to outside said apparatus.

10. The document information processing apparatus according to Claim 1, wherein said element refinement processing unit includes an element relation analyzer for associating a meaningful element with each of the meaningful words included in the structured document generated by said syntax analyzer, the meaningful element indicating a meaning of each of the meaningful words.

11. The document information processing apparatus according to Claim 10, wherein said apparatus further includes a proper name element processing unit for adding a markup to a proper name element indicating a proper name, and said element relation analyzer starts said proper name element processing unit when the meaningful element is a proper name element.

12. The document information processing apparatus according to Claim 10, wherein said apparatus further includes a time element processing unit for adding a markup to a time
5 element indicating a time, and said element relation analyzer starts said time element processing unit when the meaningful element is a time element.

13. The document information processing apparatus
10 according to Claim 12, wherein said apparatus further includes a time processing unit for computing a date corresponding to a time element, and said time element processing unit adds the date from said time processing unit, as a markup, to the time element.

15

14. The document information processing apparatus according to Claim 10, wherein said apparatus further includes a position element processing unit for adding a markup to a position element indicating a position, and said element
20 relation analyzer starts said position element processing unit when the meaningful element is a position element.

15. The document information processing apparatus according to Claim 14, wherein said apparatus further includes
25 a current position estimation unit for estimating a current position corresponding to the position element, and said position element processing unit adds the current position from said current position estimation unit, as a markup, to the position element.

30

16. The document information processing apparatus according to Claim 10, wherein said apparatus further includes an act element processing unit for adding a markup to an act element indicating an act, and said element relation analyzer
5 starts said act element processing unit when the meaningful element is an act element.